NMSim Web Server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins

Dennis M. Krüger, Aqeel Ahmed and Holger Gohlke*

Computational Pharmaceutical Chemistry Group, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, 40225 Düsseldorf, Germany

ABSTRACT

The NMSim web server implements a three-step approach for multiscale modeling of protein conformational changes. First, the protein structure is coarse-grained using the FIRST software. Second, a rigid cluster normal-mode analysis provides low-frequency normal modes. Third, these modes are used to extend the recently introduced idea of constrained geometric simulations by biasing backbone motions of the protein, whereas side chain motions are biased toward favorable rotamer states (NMSim). The generated structures are iteratively corrected regarding steric clashes and stereochemical constraint violations. The approach allows performing three simulation types: unbiased exploration of conformational space; pathway generation by a targeted simulation; and radius of gyration-guided simulation. On a data set of proteins with experimentally observed conformational changes, the NMSim approach has been shown to be a computationally efficient alternative to molecular dynamics simulations for conformational sampling of proteins. The generated conformations and pathways of conformational transitions can serve as input to docking approaches or more sophisticated sampling techniques. The web server output is a trajectory of generated conformations, Jmol representations of the coarse-graining and a subset of the trajectory and data plots of structural analyses. The NMSim webserver, accessible at http://www.nmsim.de, is free and open to all users with no login requirement.

INTRODUCTION

The intrinsic ability of biological macromolecules to switch between conformationally distinct states under native conditions leads to conformational transitions that occur on a wide range of scales, both in time and space. Such transitions are well known in the case of ligand binding to pharmacologically important proteins, e.g. HIV-1 protease, aldose reductase, FK506 binding protein, renin and DHFR (1). This kind of mutual conformational adaptation of binding partners is referred to as plasticity. Being able to understand and predict relevant conformational changes in proteins is central for an in-depth understanding of molecular recognition processes and success in structure-based ligand design (2).

In the field of computational biophysics, molecular dynamics (MD) simulation is one of the most widely applied and accurate techniques currently being used (3). Several efforts have been made to overcome the problem of restricted sampling in MD due to slow barrier crossing on the rugged energy landscape of biomacromolecules (4). However, MD simulations are still computationally expensive. Hence, there is a need to develop computational approaches that are computationally more efficient in exploring the conformational space.

Two such approaches have been shown to be promising in that respect. First, the distance geometry-based approach CONCOORD (5) generates protein conformations by satisfying distance constraints derived from a starting structure. Second, FRODA (6) generates conformations by simulating diffusive motions of flexible regions and rigid clusters of proteins by way of constrained geometric simulation. However, neither one of these approaches uses any directional guidance for sampling biologically relevant conformations; thus, both neglect valuable information related to the fact that conformational changes

---

upon ligand binding occur preferentially along lowest frequency (energy) normal modes of the unbound protein. These normal modes involve large amplitude and correlated motions (7). Normal modes can be efficiently and accurately predicted by coarse-grained normal mode approaches such as the elastic network model (ENM) (8) and the rigid cluster normal-mode analysis (RCNMA) (9).

In this study, the NMSim web server is presented, which builds upon a three-step approach introduced recently for multiscale modeling of protein conformational changes (Figure 1) (10). Initially, static properties are determined from an all-atom representation of the protein by decomposing the molecule into rigid clusters and flexible regions using the graph theoretical approach FIRST (11). In a second step, dynamical properties of the molecule are revealed by the rotations/translations of blocks approach (12) using an ENM representation of the coarse-grained protein, as implemented in the method RCNMA (9). In this step, only rigid body motions are allowed for rigid clusters, while links between them are treated as fully flexible. In the final step, termed NMSim (10), the recently introduced idea of constrained geometric simulations of diffusive motions in proteins is extended. New protein conformations are generated in that backbone motions are biased toward directions that lie in the subspace spanned by low-frequency normal modes, whereas side chain motions are biased toward attractive

FIRST
(at atomic level)

Structural rigidity

RCNMA
(at coarse-grained level)

Normal mode directions

NMSim
(at atomic level)

Conformation generation

Figure 1. Flowchart of the NMSim web service illustrating the normal mode-based geometric simulation approach.

basins derived from experimental rotamer information. The generated structures are then iteratively corrected regarding steric clashes and constraint violations. Particular attention has been given to the stereochemical accuracy of the generated conformations. This requires that backbone torsion angles are constrained to favorable regions, in addition to covalent and non-covalent bonds (hydrogen bonds and hydrophobic interactions) having proper stereochemical parameters. In total, when applied repetitively over all three steps, the procedure efficiently generates a series of conformations that lie preferentially in the subspace spanned by low-frequency normal modes.

The NMSim web service integrates these three steps and adds a layer of input and output interfaces such that trajectories of biologically relevant protein movements can be generated and analyzed in a user-friendly manner and on a reasonable timescale. The web service requires as input a PDB file or the PDB-ID of the protein of interest and a second PDB file of the target conformation in the case of a targeted simulation. The output is interactively visualized in the web browser, and a link to the results can optionally be obtained by email. The results contain the trajectory of generated conformations in PDB format, a visualization of the rigid cluster decomposition, a visualization of the generated trajectory, a graph showing the Cα atom root mean square deviations (RMSD) to the starting structure over the trajectory and a graph showing the Cα atom root mean square fluctuations (RMSF) over the trajectory. For the molecular visualizations, the web browser applet Jmol (http://jmol.sourceforge.net) is used. To the best of our knowledge, there are no other web services for normal mode-based geometric simulations that provide a similar level of integrated simulation, analysis and visualization capabilities.

MATERIALS AND METHODS

NMSim approach

The NMSim procedure starts with the analysis of structural rigidity of the input protein structure, represented as a constraint network containing covalent and non-covalent bonds, by applying FIRST (11). This yields a decomposition of the protein into rigid clusters and flexible regions in between. Subsequently, normal modes are calculated for the coarse-grained input structure by the RCNMA module. Finally, the NMSim module, first, moves the structure along directions of low-frequency normal modes and, second, generates a stereochemically valid conformation from the distorted structure. The RCNMA and NMSim modules are repetitively called during the simulation. In each RCNMA call, a new set of normal modes is calculated using the structure previously generated by NMSim as input. The RCNMA and NMSim approaches have been introduced in (9,10).

Three different types of simulations are implemented in NMSim: (i) unbiased exploration of conformational space; (ii) pathway generation by a targeted simulation; and (iii) radius of gyration-guided (ROG-guided) simulation (10). Unbiased NMSim simulation is suited for
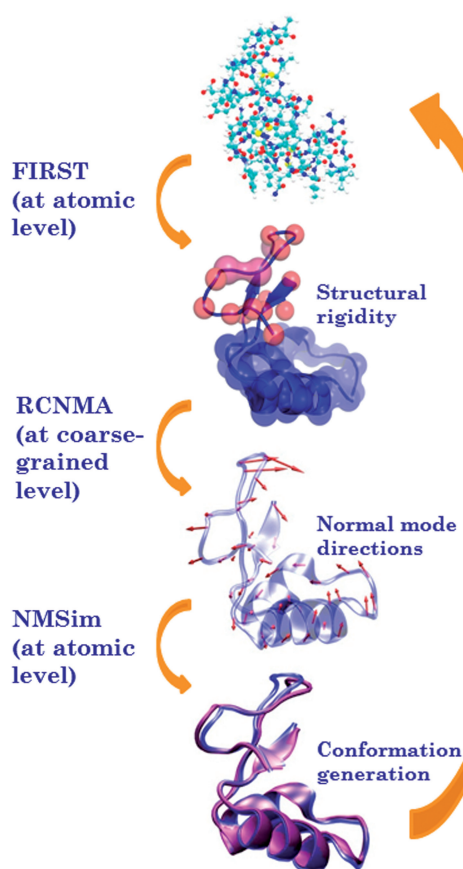
generating a conformational ensemble of a protein. This ensemble can be used for studying protein intrinsic motions or as input to other computational approaches such as docking or MD simulations. Two different parameter sets have been defined for exploring either small-scale motions (e.g. loop motions) or large-scale motions (e.g. opening and closing of domains). Targeted NMSim simulation is used for generating a pathway between a starting and a target structure. ROG-guided simulation is useful in generating ligand-bound conformations from an unbound structure. The main implementation differences in these types of simulation are described in the following sections:

Linear combination of modes in unbiased NMSim simulation

For an unbiased NMSim simulation, a random linear combination of normal modes is used [Equation (1)]:

$$\vec{V}_i = \sum_{k=7}^{m} \frac{O^k}{\omega^k} \vec{P}_{i(j)}^{k} \tag{1}$$

A coefficient in the linear combination is defined as the ratio of a uniformly distributed random number $O^k \in [-1,1]$ and a factor $\omega^k = \sqrt{\Lambda^k}$. Here, Λ^k is the eigenvalue of normal mode k. The direction of displacement of atom i of residue j is given by the sum over vectors $\vec{P}_{i(j)}^{k}$; $\vec{P}_{i(j)}^{k}$ are derived from Cα atom-based normal modes according to Equation (1) in (10). All $\vec{P}_{i(j)}^{k}$ vectors belonging to the m lowest frequency normal modes (except for the first six zero-frequency modes) are considered in the linear combination. The parameter m can be specified by the user.

Linear combination of modes in targeted NMSim simulation

In a targeted NMSim simulation, a conformational change vector $\vec{r} = \vec{r}_c - \vec{r}_o$ is used to guide the trajectory toward a target structure $\vec{r}_c$, starting from a structure $\vec{r}_o$. The vectors $\vec{r}_c$ and $\vec{r}_o$ are the Cα atomic coordinates of the two conformations. A coefficient Z^k [Equation (2)] is now obtained as the projection of the conformational change vector $\Delta\vec{r}$ onto the normal mode vector $\vec{C}^k$.

$$Z^k = \Delta\vec{r} \cdot \vec{C}^k \tag{2}$$

Z^k is used to bias the linear combination of displacement vectors of atom i of residue j according to Equation (3).

$$\vec{V}_i = \sum_{k=7}^{m} Z^k \vec{P}_{i(j)}^{k} \tag{3}$$

ROG-guided NMSim simulation

The search for a ligand bound conformation of a protein can be drastically improved if structural characteristics of the complex are incorporated to guide the search. In the case of large-scale conformational changes, e.g. domain closures in proteins upon ligand binding, it is well known that the compactness of the protein structure increases upon binding (13). The ROG R_g [Equation (4)] is an appropriate measure to describe the compactness of a protein:

$$R_g^2 = \frac{1}{n} \sum_{i=1}^{n} (\vec{r}_i - \vec{r}_c)^2 \tag{4}$$

where $\vec{R}_c$ is the center of geometry of n Cα atoms and $\vec{r}_i$ is the atomic position of atom i. Here, only Cα atoms are considered.

In an ROG-guided NMSim simulation, the trajectory is tailored toward the bound structure by selecting the pathway that leads to a decrease in R_g. We note that R_g is not used for enforcing the generation of a more compact protein structure by influencing the structure distortion step: conformations are still generated by structure distortion along directions of random linear combinations of low-frequency normal modes [Equation (1)] and subsequent structure correction. Rather, the guiding occurs because out of several (by default, three) such conformations generated by repeated NMSim cycles, the one with the lowest R_g is selected as a starting point for further trajectory exploration in the next simulation cycle.

Structure correction

A distorted structure is efficiently corrected using a geometry-based constraints correction approach. For this, the network of constraints from the rigidity analysis is used where covalent and non-covalent bonds are considered as constraints. In addition, constraints for φ/ψ backbone torsion angles are introduced, which are derived from favorable regions on a Ramachandran map. For χ-angles of side chains, a knowledge-based approach is applied by forcing side chains to move into the closest favorable rotamer state during the structure correction. Finally, chirality and planarity constraints are considered for backbone and side chains, and steric clashes between atoms are removed. For further details see (10).

Description of the web server

Input

The NMSim web service submission page is shown in Figure 2. The NMSim web service requires either a structure of a protein provided as a PDB file or a PDB-ID (in which case the PDB file will be downloaded from the RCSB repository), selection of a simulation type and input of a given security code to prevent misuse of the web service. In the case of a targeted simulation, two PDB files are required, with the first one being the starting structure and the second one the target structure of the protein. According to the selected type of simulation, default parameters will be provided, which have turned out to provide reasonable results in a validation study (10). Optionally, an email address can be provided, in which case a link to the results will be sent to that address. In either case, a link to a result page is provided after job submission for monitoring the progress of the computations and viewing the results in

Figure 2. Screenshot of the NMSim web server submission page.

the web browser. The results will also be stored on the server for 10 days. Furthermore, it is possible to test the level of coarse-graining of the structure without starting a simulation by applying the 'Preview RCD' button. The user is encouraged to make use of this feature as providing meaningful parameters for the rigid cluster decomposition is crucial for the simulation success. As such, a structure can be made more floppy by lowering the values for 'E-cutoff for H-bonds' and/or 'Hydrophobic cutoff'. Note that these preview results are only provided for viewing in the web browser.

Some restrictions apply to the PDB input file. (i) Hydrogen atoms are not required because they are automatically added before a simulation run using the program Reduce (14). Accordingly, for amino acid side chains, a standard protonation state is assumed, i.e. Asp and Glu are treated as deprotonated and Arg, Lys and His

as protonated. (ii) Alternative side chain conformations are not allowed. If present in the PDB file, an error message will be issued. Furthermore, it is crucial that the protein structure is complete, i.e. no missing residues are allowed. In case of a missing side chain, atoms will be automatically added by the leap program of the AMBER suite of programs (15). Finally, if a targeted simulation is requested, starting and target protein structures must be identical except for the coordinates. If this is not given, an error message will be issued.

Output and representation of results

A typical simulation run of the NMSim web service takes a few hours. After a simulation is started, the estimated overall runtime and the progression of the simulation are given on the results page. Upon completion of a job, the results are presented in the web browser and, if an address

is provided by the user, an email is sent with a link to the results too. The first part of the results page contains a summary of input parameters that were provided by the user and a download link to the trajectory file. The trajectory file contains all sampled conformations and is compressed with gzip. This file can be used to animate protein motions as in a movie, e.g. with PyMol (http://pymol.sourceforge.net) or VMD (http://www.ks.uiuc.edu/Research/vmd/), or to extract structures for further modeling purposes, e.g. molecular docking experiments.

The second part of the results page contains two Jmol representations: (i) the rigid cluster decomposition (Figure 3A) and (ii) a subset of 20 conformations extracted at equally spaced intervals from the NMSim trajectory (Figure 3B). The rigid cluster decomposition shows all

atoms of the protein that are part of one rigid cluster with one color, with the largest cluster being depicted in blue. The subset of conformations gives the user the possibility to animate protein movements directly in the web browser within the JMol Plugin using 'Mouse → Right click → Animation → Play'. The starting structure is depicted in blue and the ending structure in green. All other conformations corresponding to the secondary structure elements are depicted with following colors: helices in magenta, beta sheets in orange and loops in white.

The third part contains two graphs that provide information about the conformational changes of the protein during the simulation (Figure 4): (i) a graph showing the Cα atom RMSD to the starting structure over the trajectory and (ii) a graph showing the Cα atom RMSF over the trajectory. Both the RMSD and RMSF are computed after superimposing the generated conformations onto the starting structure with respect to the Cα atoms. A file with the raw data can be downloaded by a link given under each graphical representation for further evaluation.
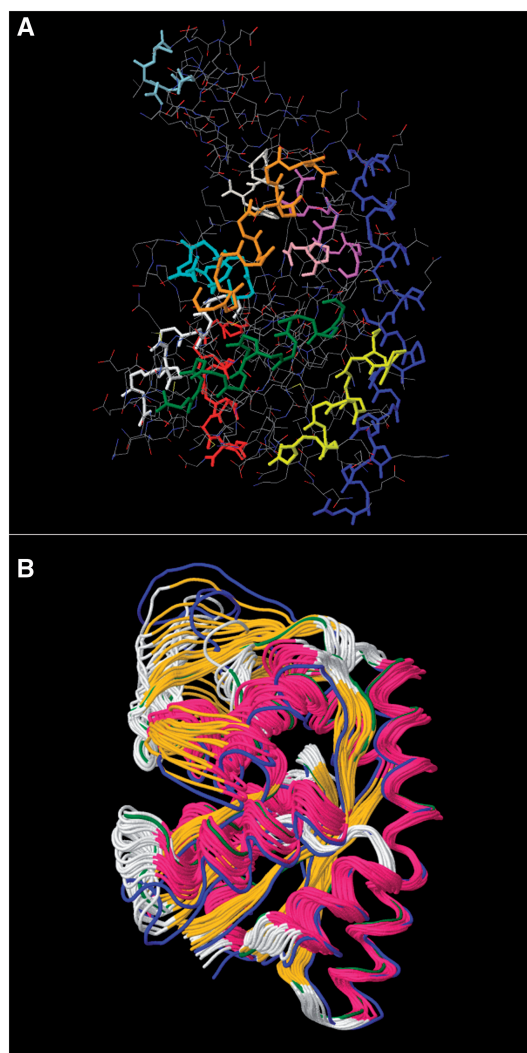

Figure 3. (**A**) Screenshot of the color-coded representation of the rigid cluster decomposition of adenylate kinase (PDB code: 4AKE). The largest rigid cluster is shown in blue. (**B**) Screenshot of 20 conformations extracted at equally spaced intervals from the trajectory of an ROG-guided NMSim run of adenylate kinase (PDB code: 4AKE). The starting structure is depicted in blue, the ending structure in green and all other conformations corresponding to the secondary structure elements.
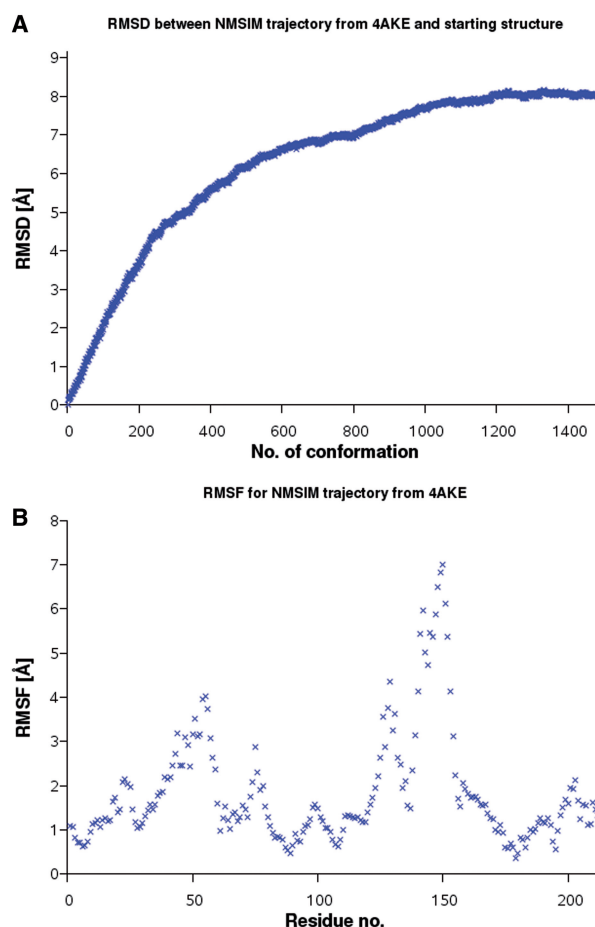

Figure 4. (**A**) Graph showing the Cα atom RMSD to the starting structure over the trajectory obtained for an ROG-guided NMSim run of adenylate kinase (PDB code: 4AKE). (**B**) Graph showing the Cα atom RMSF over the trajectory obtained for an ROG-guided NMSim run of adenylate kinase (PDB code: 4AKE).

If only the 'Preview RCD' option is used for testing the level of coarse graining of the input structure, a typical run takes a few minutes. Upon completion of a job, the results are presented in the web browser in an additional window. The results page contains the input parameters and a Jmol representation of the rigid cluster decomposition.

Implementation

The NMSim web server has been implemented in Python. The RMSD and RMSF are calculated with the ptraj program included in AmberTools (15). The RCNMA and NMSim routines have been implemented in Fortran. Given the low computational demand of our approach, up to four submitted jobs can be run in parallel at present.

Application to adenylate kinase as a test case

In a recent validation study of the NMSim approach (10), conformational variabilities were reproduced very well for four of five proteins with experimentally observed domain motions, with correlation coefficients $r > 0.70$. In seven of eight cases, including domain and loop motions, NMSim simulations starting from unbound structures were able to sample conformations that are similar (RMSD = 1.0– 3.1 Å) to ligand-bound conformations. These results showed in general that incorporating directional information about collective motions into a constrained geometric simulation-based approach allows for a thorough sampling of the biologically relevant conformational space and provides a computationally efficient alternative to MD simulations for conformation generation.

As an explicit example, we describe results obtained for adenylate kinase (ADK) using the NMSim web server in more detail now. These results are also provided as a sample run on the web server. ADK was chosen for this because it is a well-studied protein in terms of catalytic mechanism and conformational flexibility and has been used as a test case in other computational studies. We intended to sample structures that are close to a ligand-bound ('closed') conformation (PDB code: 1AKE) starting from an input structure in the unbound conformation (PDB code: 4AKE). Therefore, ROG-guided simulation was performed using default parameters. The ROG-guided simulation is based on the assumption that ligand binding usually results in a more compact protein conformation due to domain or loop closure. Thus, no *a priori* knowledge about the closed structure has been applied to foster the transition to the closed structure. The coarse-graining of the protein structure is depicted in Figure 3A. Figure 3B illustrates the extent by which ROG-guided NMSim simulations were successful in sampling domain motions toward the closed structure. The large-scale conformational change of the LID domain is well described; however, the NMPbind domain moves only halfway toward the closed structure, in agreement with the suggestion that its closing follows a ligand-induced mechanism (16). The conformation that is most similar to the experimentally determined closed structure is shown in green and has an RMSD of 2.4 Å to that structure; in contrast, the starting structure (blue) has an RMSD of 7.15 Å to the closed conformation.

Notably, conformational variabilities have been reproduced very well for this example with a correlation coefficient $r = 0.92$ compared to experimentally observed conformational changes (10).

CONCLUSIONS

The development of the NMSim web server was motivated by a growing demand for comprehensive tools to efficiently predict conformational changes in biomacromolecules. The NMSim web server integrates a novel three-step approach, termed NMSim, for multiscale modeling of protein conformational changes that incorporates information about preferred directions of protein motions into a geometric simulation algorithm. This procedure efficiently generates a series of stereochemically allowed conformations. Incorporating directional information distinguishes the NMSim approach from other widely used geometry-based simulation approaches. The NMSim web server allows performing three simulation types: (i) unbiased exploration of the conformational space; (ii) pathway generation by a targeted simulation; and (iii) ROG-guided simulation to foster the compaction of a protein structure upon ligand binding. We note that in the latter case, no experimental ROG is required as input. The NMSIM approach has already been successful validated: when applied to a data set of proteins where conformational changes have been observed experimentally, either as domain motions or motions of functionally important loops, experimental conformational variabilities were reproduced very well by the NMSim simulations (10). The NMSim web server is primarily intended for generating trajectories of protein conformations online. A user-friendly interface, minimal demands on input information and a detailed output as well as an embedded visualization capability make this web server potentially useful for users without prior knowledge of structural bioinformatics analyses. Overall, we expect the NMSim web server to be a valuable tool for the sampling of protein conformations that can then serve as input to docking approaches or as starting points for more sophisticated sampling techniques.

REFERENCES

1. Lin,J.H., Perryman,A.L., Schames,J.R. and McCammon,J.A. (2002) Computational drug design accommodating receptor

flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.*, **124**, 5632–5633.

2. Cavasotto,C.N. and Abagyan,R.A. (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.*, **337**, 209–225.

3. McCammon,J.A., Gelin,B.R. and Karplus,M. (1977) Dynamics of folded proteins. *Nature*, **267**, 585–590.

4. Cheatham,T.E. III and Young,M.A. (2000) Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, **56**, 232–256.

5. Seeliger,D., Haas,J. and de Groot,B.L. (2007) Geometry-based sampling of conformational transitions in proteins. *Structure*, **15**, 1482–1492.

6. Wells,S., Menor,S., Hespenheide,B. and Thorpe,M.F. (2005) Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.*, **2**, S127–S136.

7. Tama,F. and Sanejouand,Y.H. (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng.*, **14**, 1–6.

8. Hinsen,K. (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins*, **33**, 417–429.

9. Ahmed,A. and Gohlke,H. (2006) Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins*, **63**, 1038–1051.

10. Ahmed,A., Rippmann,F., Barnickel,G. and Gohlke,H. A normal mode-based geometric simulation approach for exploring biologically relevant conformational transitions in proteins. *J. Chem. Inf. Model.*, **51**, 1604–1622.

11. Jacobs,D.J., Rader,A.J., Kuhn,L.A. and Thorpe,M.F. (2001) Protein flexibility predictions using graph theory. *Proteins*, **44**, 150–165.

12. Durand,P., Trinquier,G. and Sanejouand,Y.-H. (1994) A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers*, **34**, 759–771.

13. Egea,P.F., Rochel,N., Birck,C., Vachette,P., Timmins,P.A. and Moras,D. (2001) Effects of ligand binding on the association properties and conformation in solution of retinoic acid receptors RXR and RAR. *J. Mol. Biol.*, **307**, 557–576.

14. Word,J.M., Lovell,S.C., Richardson,J.S. and Richardson,D.C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, **285**, 1735–1747.

15. Case,D.A., Cheatham,T.E. III, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comp. Chem.*, **26**, 1668–1688.

16. Whitford,P.C., Gosavi,S. and Onuchic,J.N. (2008) Conformational transitions in adenylate kinase. Allosteric communication reduces misligation. *J. Biol. Chem.*, **283**, 2042–2048.